決定理論による定式化における直交計画を用いたブール関数の学習に関する一考察

浮 田 善 文

1．まえがき

質問からの学習とは学習者が能動的に選択する例から関数を学習する学習パラダイムであり，これまで多くの研究が行われている．Angluin[1]はk-termDNF，k-clauseCNF等の関数クラスに対し６種類の質問（所属性，等価性，部分性，包含性，非交差性，全体性）のうちどの種類の質問を用いれば多項式時間でEXACT学習[1]が可能か否かを示している．また，浮田ら[15]は，関数を実数体上の直交基底によって表現したときに高次の成分を含まない関数クラスに対し所属性質問を用い多項式時間でEXACT学習可能であることを示している．このように質問からの学習に関する研究の多くは，関数クラスに対しある効率（計算量）である学習成功基準を達成する学習が可能か否かを判定する研究であった．

ここで，質問回数が制限されている場合，質問する入力の集合（質問集合）によって真の関数を正しく学習する確率が変わるため，どの質問集合を選ぶかが重要になる．しかし従来の研究の多くは計算論的に考察されており，学習成功基準と学習効率基準を満たす質問戦略が存在するか否かが明らかにされるのみでどの質問集合が良いかまでは評価することができなかった．

さて，質問からの学習と類似の研究分野に統計学の実験計画法[9, 13]があげられる．この研究分野では決定理論に基づいた定式化がなされ，同じ実験回数の実験計画の中で最適（各係数の不偏推定量の分散の最大値が最小）な実験計画である直交計画に関する研究が多く行われている．

浮田ら[14]は，質問からの学習問題を決定理論的に定式化することで，質問集合の良さを評価可能な評価方法を提案した．そして，質問回数（学習効率基準）を固定したもとで平均決定誤り率（学習成功基準）を最小とする決定と質問集合を求める分枝限定アルゴリズム[1]が提案されている．ここでは，各ブール関数に事前確率[2]を付置した一般的な事前確率分布を対象としているが，アルゴリズムの計算量はブール変数の個数 n と質問回数 K の指数オーダであるため，n や K が大きい場合には実時間での計算は困難であった．

そこで本論文では，文献[14]と同様に決定理論による定式化における質問からの学習を扱う．ただし，一般的な事前確率分布を対象とするのではなく，出力に影響を与えるブール変数の個数が大きいブール関数ほどそのブール関数の事前確率は小さくなる事前確率分布を対象とした質問学習を扱う．この条件を満たす場合，事前確率の高いブール関数集合は文献[15]で対象とする関数クラスとなる．そこで，[15]と同様に質問集合として直交計画を用いるアルゴリズムを提案する．次に，このアルゴリズムにおいて，質問後にブール関数を決定するのに必要な計算量は n と K の多項式オーダであることを示す．さらに，提案アルゴリズムは質問結果に対してベイズ決定を行うことを示す．最後に，提案アルゴリズムを用いたときの平均決定誤り率を示し，直交計画が最適質問集合（ベイズ決定を行うときの平均決定誤り率を最小にする質問集合）となる条件を与える．

以下で本論文の流れを説明しておく．まず2で準備として，本論文で必要となる定義と従来研究の結果を与えておく．3では，決定理論に基づく質問学習のモデル化を与え，平均決定誤り率を定義する．そして，平均決

定誤り率を最小化する決定関数（ベイズ決定）と最適質問集合を与えておく．4では，本論文で扱う質問学習の問題設定を与える．5では，質問集合に直交計画を用いたアルゴリズムを提案する．6では，提案アルゴリズムがベイズ決定を行うことを示す．7では，提案アルゴリズムを用いたときの平均決定誤り率を示し，直交計画が最適質問集合となる条件を与える．8では，アルゴリズムの実行例を示す．最後に9でまとめを行う．

2．準備

2.1　記号の定義

始めに，必要な用語を与えておく．長さ n のベクトルを $\underline{a} = (a_1, a_2, \cdots, a_n)$，$\underline{b} = (b_1, b_2, \cdots, b_n)$ と表し，$\underline{a}, \underline{b} \in \{0, 1\}^n$ とする．ここで，ベクトル間の加法＋を，

$$\underline{a} + \underline{b} = (a_1 \oplus b_1, a_2 \oplus b_2, \cdots, a_n \oplus b_n), \quad \cdots\cdots\cdots\cdots (1)$$

とする．ただし，⊕は排他的論理和である．ベクトル間の内積・を，

$$\underline{a} \cdot \underline{b}^T = a_1b_1 \oplus a_2b_2 \oplus \cdots \oplus a_nb_n, \quad \cdots\cdots\cdots\cdots (2)$$

とする．ただし，記号 T は転置を表す．

ベクトル $\underline{a}$ のハミング重み $w(\underline{a})$ を，$w(\underline{a}) = \Sigma_{i=1}^{n} a_i$ とする．また，全ての要素が 0 であるベクトルを $\underline{0}$ とする[2]．

定義1　1次独立

$u_i \in \{0, 1\}, (1 \leq i \leq l)$ とする．$u_1\underline{c_1} + u_2\underline{c_2} + \cdots + u_l\underline{c_l} = 0$ が成立するのは，$u_1 = u_2 = \cdots = u_l = 0$ のみであるとき，$\{\underline{c_1}, \underline{c_2}, \cdots, \underline{c_l}\}$ は1次独立という．　□

定義２　行列 G

$k \times n$ 行列 G を次式で定義する．

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \vdots & & \vdots \\ g_{k1} & \cdots & g_{kn} \end{bmatrix}, \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (3)$$

ただし，$g_{ij} \in \{0, 1\}, (1 \leq i \leq k, 1 \leq j \leq n)$ である．

G の第 i 行を $\underline{g_{i\cdot}} = [g_{i1}\ g_{i2} \cdots g_{in}]$, G の第 j 列を $\underline{g_{\cdot j}} = [g_{1j}\ g_{2j} \cdots g_{kj}]^T$ と表記する．行列 G の行ベクトルは１次独立，すなわち，$\{\underline{g_{1\cdot}}, \underline{g_{2\cdot}}, \cdots, \underline{g_{k\cdot}}\}$ は１次独立とする．□

行列 G の階数を $rank(G)$ と表記する．

$v(\underline{a}) = \{i | a_i \neq 0, 1 \leq i \leq n\}$, $v(\underline{a}, \underline{b}) = \{i | a_i \neq 0, 1 \leq i \leq n\} \triangle \{i | b_i \neq 0, 1 \leq i \leq n\}$ とする．ただし，$S_1 \triangle S_2$ は，集合 S_1, S_2 の対称差を表す．

2. 2　直交基底による関数の表現

$\underline{x} = (x_1, x_2, \cdots, x_n) \in \{0, 1\}^n$ とする．任意のベクトル $\underline{a} = (a_1, a_2, \cdots, a_n) \in \{0, 1\}^n$ に対して以下に定義される関数 $\mathcal{X}_{\underline{a}} : \{0, 1\}^n \to \{+1, -1\}$ を，基底関数と呼ぶ．

$$\mathcal{X}_{\underline{a}}(\underline{x}) = (-1)^{\underline{a} \cdot \underline{x}^T}. \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (4)$$

このとき，基底関数のクラス $\{\mathcal{X}_{\underline{a}} | \underline{a} \in \{0, 1\}^n\}$ は正規直交系をなす．すなわち，

$$\frac{1}{2^n} \sum_{\underline{x} \in \{0,1\}^n} \mathcal{X}_{\underline{a}}(\underline{x}) \mathcal{X}_{\underline{b}}(\underline{x}) \begin{cases} 1; \underline{a} = \underline{b}, \\ 0; \underline{a} \neq \underline{b}, \end{cases} \quad \cdots\cdots\cdots\cdots\cdots \quad (5)$$

が成立する．このことから，任意の実数値関数 f: $\{0, 1\}^n \to \mathcal{R}$ は次式により，基底関数の線形結合として一意に表すことができる．

$$f(\underline{x}) = \sum_{\underline{a}\in\{0,1\}^n} f_{\underline{a}}\,\mathcal{X}_{\underline{a}}(\underline{x}). \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (6)$$

すべての関数値 $f(\underline{x})$, ($\underline{x} \in \{0, 1\}^n$) が与えられる場合，次式により係数 $f_{\underline{a}}$ を計算することができる．

$$f_{\underline{a}} = \frac{1}{2^n} \sum_{\underline{x}\in\{0,1\}^n} f(\underline{x})\,\mathcal{X}_{\underline{a}}(\underline{x}). \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (7)$$

2.3 直交計画を用いた実数値関数の学習 [15]

まず，関数クラスを定義するのに必要な仮定を与えておく．

仮定 1

集合 A, ($A \subseteq \{0, 1\}^n$) は，$\underline{a} \in A$ ならば，任意の $\underline{b}$, ($\underline{b} \in \{0, 1\}^n$, $\underline{b} \sqsubseteq \underline{a}$) に対して $\underline{b} \in A$ が成立する．ただし，$\underline{b} \sqsubseteq \underline{a}$ は，すべての i, ($1 \leq i \leq n$) において $b_i \leq a_i$ であることを表す． □

例 1　集合 A の例

$A = \{0000, 1000, 0100, 1100, 0010, 1010, 0110, 0001\}$. □

定義 3　実数値関数クラス $\mathcal{F}_A^{\mathcal{R}}$

仮定 1 を満たす，ある A を用いて次式で表現可能な実数値関数の集合を $\mathcal{F}_A^{\mathcal{R}}$ とする．

$$f(\underline{x}) = \sum_{\underline{a}\in A} f_{\underline{a}}\,\mathcal{X}_{\underline{a}}(\underline{x}). \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (8)$$

ただし，実数値関数の定義域と値域は $\underline{x} \in \{0, 1\}^n$, $f(\underline{x}) \in \mathcal{R}$ である． □

次に直交計画を定義しておく．

定義 4　直交計画 X_A

$\forall \underline{a}, \underline{b} \in A$ に対して，$\{\underline{g}_{\cdot j} | j \in v(\underline{a}, \underline{b})\}$ が 1 次独立であり，かつ，行ベク

トルが１次独立となる行列を G_A とする．なお，G_A の行数は集合 A に依存するため，k_A で表す．

このとき，直交計画 X_A は，$k_A \times n$ 行列 G_A を用い，次式で与えられる．

$$X_A = \{\underline{x} \mid \underline{x} = \underline{r} G_A, \underline{r} \in \{0, 1\}^{k_A}\}. \quad \cdots\cdots\cdots\cdots\cdots \quad (9)$$

ここで，$|X_A| = 2^{k_A}$ である．　□

直交計画の例を例２に載せる．

例２　直交計画

$n = 4, A = \{0000, 1000, 0100, 1100, 0010, 1010, 0110, 0001\}$ の場合を考える．このとき，定義４の条件を満たす G_A として，

$$G_A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \cdots\cdots\cdots\cdots\cdots\cdots \quad (10)$$

があげられる．このとき，直交計画 $X_A = \{0000, 1100, 1010, 0110, 1001, 0101, 0011, 1111\}$ となる．

なお，G_A は有限射影幾何を用いたアルゴリズム[13]などにより求めることができる．　□

真の実数値関数を $f^*(\underline{x}) = \Sigma_{\underline{a}} f^*_{\underline{a}} \mathcal{X}_{\underline{a}}(\underline{x})$ とする．直交計画を用いた実数値関数の学習に関する定理を以下に載せる．

定理１ [15]

$f^*(\underline{x}) \in \mathcal{F}^{\mathcal{R}}_A$ であれば，係数 $f^*_{\underline{a}}$ は次式により計算される．

$$f^*_{\underline{a}} = \frac{1}{|X_A|} \sum_{\underline{x} \in X_A} f^*(\underline{x}) \mathcal{X}_{\underline{a}}(\underline{x}). \quad \cdots\cdots\cdots\cdots\cdots \quad (11)$$

□

3　決定理論に基づく質問学習のモデル化

この章では，決定理論に基づく質問学習のモデル化[14]について説明する．

3.1　決定理論に基づく質問学習

本論文では入出力が有限の離散値でありブール関数の個数が有限の場合の質問学習を決定理論的に定式化する．まず質問学習の枠組みを定義する．

定義5　質問学習の枠組み

ブール関数を $f \in \mathcal{F}$, ブール関数の入力を $\underline{x} \in \mathcal{X} = \{0, 1\}^n$, 出力を $f(\underline{x}) \in \mathcal{Y} = \{+1, -1\}$ と表す．

次に，質問する K 個の入力列を $X = (\underline{x_1}, \underline{x_2}, \cdots, \underline{x_K}) \in \mathcal{X}^K$, 入力列に対応して得られる K 個の出力列を $f(X) = (f(\underline{x_1}), f(\underline{x_2}), \cdots, f(\underline{x_K})) \in \mathcal{Y}^K$ で表し，これらの対 $X, f(X)$ をデータと呼ぶ．なお，X は集合 $X = \{\underline{x_1}, \underline{x_2}, \cdots, \underline{x_K}\}$ としても用いる．質問学習とは，質問により得られたデータ $(\underline{x_i}, f(\underline{x_i}))_{i=1}^{K}$ から未知である真のブール関数 f を学習する問題である．

決定理論に基づく質問学習ではブール関数の事前確率が与えられる．ブール関数 f' の事前確率が $p(f')$ であるとすると，データ $X, f(X)$ を観測した後の f' の事後確率は，次式で計算される．

$$p(f' | X, f(X)) = \begin{cases} \dfrac{p(f')}{\sum_{g \in F_{X, f(X)}} p(g)}, & f' \in F_{X, f(X)}; \\ 0, & f' \notin F_{X, f(X)}. \end{cases} \quad \cdots\cdots\cdots\cdots \quad (12)$$

ただし，$F_{X, f(X)}$ は $X, f(X)$ に無矛盾なブール関数の集合，すなわち，$F_{X, f(X)} = \{g | g(X) = f(X), g \in \mathcal{F}\}$ である．　□

質問学習におけるデータ選択基準の１つとして，決定の期待損失を最小にするようにデータを選ぶことが考えられる．この基準は，ベイズ決定理論の枠組みで定式化することができる．まず，ベイズ決定理論の考え方を説明しておく．個々の決定の良さは，決定の目的に依存する，という状況を考える．質問後に決定 d を行ったときに真の値が α である場合の損失が損失関数 $L(d, \alpha) \in [0, \infty)$ で計ることができるものとする．ここで，d は決定関数と呼ばれ，どの決定関数を用いるかは推論の目的に依存する．

3.1.1　決定関数

推論の目的による分類と決定関数の関係は文献[8]に詳しく述べられている．この節では，代表的な推論である帰納推論とデータからの予測を目的としたときの決定関数を以下に示す．

データ X, $f(X)$ が与えられたもとで，ブール関数 f を決定する推論は帰納推論と呼ばれ，以下の決定関数を用いる．

$$d = \hat{f}_{X, f(X)} \in \mathcal{F}. \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (13)$$

また，X, $f(X)$ と $\underline{x}'$ が与えられたもとで，$f(\underline{x}')$ を決定する推論はデータからの予測と呼ばれ，以下の決定関数を用いる．

$$d = \widehat{f(\underline{x}')}_{X, f(X)} \in \{+1, -1\}. \cdots\cdots\cdots\cdots\cdots \quad (14)$$

3.1.2　損失関数

損失関数は様々なものが考えられる[2]．一般に，損失は決定 d と真の値との間のなんらかの距離を用いて評価される．よく用いられる距離として次式があげられる．

$$l(\alpha, \beta) = \begin{cases} 0, & \alpha = \beta; \\ 1, & \alpha \neq \beta. \end{cases} \quad \cdots\cdots\cdots\cdots\cdots \quad (15)$$

例3　ブール関数 f を決定する推論を行うとき，以下の損失関数が用いられる．

$$L_1 = l(\hat{f}_{X,f(X)}, f). \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (16)$$

式(16)は，決定したブール関数が正しい場合には損失0，誤った場合には損失1が生じることを示している．

データからの予測を行うとき，以下の損失関数が用いられる．

$$L_2 = \sum_{\underline{x}' \in \mathcal{X}} l(\widehat{f(\underline{x}')}_{X,f(X)}, f(\underline{x}'))p(\underline{x}'). \quad \cdots\cdots\cdots\cdots \quad (17)$$

ここで，$p(\underline{x}')$ は X 上に仮定された確率分布に従って要素 $\underline{x}'$ が取り出される確率であり，この確率は既知とする．　□

3.1.3　評価基準

決定理論においては損失関数を最小化する決定関数を求めることが中心的課題となるが，初期の学習問題では学習が成功した基準を考え，ある関数クラスに対しそれを達成しうるか，つまり，学習可能かを判断することが主な関心事となっていた．例えば，極限同定[4]の基準は前述の損失 L_1 が0に収束することを学習成功の基準としている．また，PAC学習[19]では前述の損失 L_2 が0に確率収束する速さを学習成功の評価としている[8]．

さて，もし真のブール関数 f が既知であれば決定関数の良さは損失関数により計ることができる．しかし，実際には真のブール関数 f は未知であるため，どの f に対しても損失関数を最小とする決定関数が存在するとは限らない．そこで決定理論では，損失関数を最小化する代表的視点として，ベイズ基準，maximin基準，minimax基準[3]があげられるが，本節ではPaass[10]と同様に，ベイズ基準により損失関数を評価する．ベイズ基準ではベイズ損失関数を最小化する決定が最適な決定となる．例えば，損失関数 L_1 のベイズ損失関数は以下のように定義される．

$$BR_1(X, \hat{f}_{X,\mathcal{Y}^K}, p(f)) = \sum_{f \in \mathcal{F}} l(\hat{f}_{X,f(X)}, f)p(f). \quad \cdots\cdots \quad (18)$$

式(18)は平均決定誤り率となる．また，損失関数 L_2 を用いたときのベイズ損失は平均予測誤り率となる．本論文では平均決定誤り率によって評価するため，以後，損失関数はすべて L_1 を用いる．

3.2　ベイズ決定

ベイズ損失を最小化する決定関数はベイズ決定関数と呼ばれ，以下で定義される．

定義6　ベイズ決定関数

$$\hat{f}^{opt}_{X,\mathcal{Y}^K} = \arg\min_{\hat{f}_{X,\mathcal{Y}^K}} BR_1(X, \hat{f}_{X,\mathcal{Y}^K}, p(f)). \cdots\cdots\cdots\cdots \quad (19)$$

□

補題1

ベイズ決定関数 $\hat{f}^{opt}_{X,\mathcal{Y}^K}$ は，質問結果 $Y \in \mathcal{Y}^K$ ごとに次式で与えられる．

$$\hat{f}^{opt}_{X,\mathcal{Y}^K} = \hat{f}^{(1)}_{X,Y}. \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \quad (20)$$

ただし，$\hat{f}^{(1)}_{X,Y} = \arg\max_{f \in F_{X,Y}} p(f)$である．

（証明）　式(18)で与えられる平均決定誤り率は以下のように書き換えることができる．

$$\sum_{f \in \mathcal{F}} l(\hat{f}_{X,f(X)}, f)p(f)$$

$$= \sum_{Y \in \mathcal{Y}^K} \sum_{f \in F_{X,Y}} l(\hat{f}_{X,Y}, f)p(f). \cdots\cdots\cdots\cdots\cdots \quad (21)$$

ここで，平均決定誤り率を最小化するには，各 Y に対し $\sum_{f \in F_{X,Y}} l(\hat{f}_{X,Y}, f)p(f)$ を最小化すればよい．損失関数の定義より，次

式が成立する．

$$\sum_{f \in F_{X,Y}} l(\hat{f}_{X,Y}, f)p(f) =$$

$$\begin{cases} \Sigma_{f \in F_{X,Y} \setminus \{\hat{f}_{X,Y}\}} p(f), & \hat{f}_{X,Y} \in F_{X,Y}; \\ \Sigma_{f \in F_{X,Y}} p(f), & \hat{f}_{X,Y} \notin F_{X,Y}. \end{cases} \quad \cdots\cdots\cdots\cdots \quad (22)$$

式(22)より，$\Sigma_{f \in F_{X,Y}} l(\hat{f}_{X,Y}, f)p(f)$ を最小化する $\hat{f}_{X,Y}$ は $f^{(1)}_{X,\,Y}$ であることが分かる． □

一般には，損失関数によらず損失関数の事後分布上での期待値（期待損失）を最小化する決定がベイズ決定となる[8]．

補題２

ベイズ決定を行うときの平均決定誤り率は次式で与えられる．

$$BR_1(X, \hat{f}^{opt}_{X,\,\mathcal{Y}^K}, p(f)) = 1 - \sum_{Y \in \mathcal{Y}^K} p(\hat{f}^{(1)}_{X,\,Y}). \quad \cdots\cdots\cdots \quad (23)$$

（証明） 式(21),(22)を用いると，平均決定誤り率は以下のように式展開できる．

$$\begin{aligned} & BR_1(X, \hat{f}^{opt}_{X,\,\mathcal{Y}^K}, p(f)) \\ &= \sum_{Y \in \mathcal{Y}^K} \sum_{f \in F_{X,Y}} l(\hat{f}^{(1)}_{X,\,Y}, f)p(f) \\ &= \sum_{Y \in \mathcal{Y}^K} \sum_{f \in F_{X,Y} \setminus \{\hat{f}^{(1)}_{X,Y}\}} p(f) \\ &= 1 - \sum_{Y \in \mathcal{Y}^K} p(\hat{f}^{(1)}_{X,\,Y}). \end{aligned} \quad \cdots\cdots\cdots\cdots\cdots \quad (24)$$

以上より，補題が証明される． □

例を用い，質問する入力集合 X により平均決定誤り率がどのように変

わるかを示す.

例4　ブール変数の個数は $n=2$, 質問回数は $K=1$ とする. ここで, ブール関数の事前確率が $p(x_1x_2)=0.4$, $p(x_1 \vee x_2)=0.3$, $p(\bar{x}_1\bar{x}_2)=0.2$, $p(\bar{x}_1 \vee \bar{x}_2)=0.1$ で与えられる場合を考える. ただし, $\bar{x}$ は x の否定を表す.

- X = (11) のとき

質問結果として, $Y=(+1), (-1)$ の2通りが考えられ, $F_{(11),(+1)} = \{x_1x_2, x_1 \vee x_2\}$, $F_{(11),(-1)} = \{\bar{x}_1\bar{x}_2, \bar{x}_1 \vee \bar{x}_2\}$ である. このとき, 平均決定誤り率は $BR_1 = 1 - p(x_1x_2) - p(\bar{x}_1\bar{x}_2) = 1 - 0.4 - 0.2 = 0.4$ である.

- X = (01) のとき

質問結果として, $Y=(+1), (-1)$ の2通りが考えられ, $F_{(01),(+1)} = \{x_1 \vee x_2, \ \bar{x}_1 \vee \bar{x}_2\}$, $F_{(01),(-1)} = \{x_1x_2, \bar{x}_1\bar{x}_2\}$ である. このとき, 平均決定誤り率は $BR_1 = 1 - p(x_1 \vee x_2) - p(x_1x_2) = 1 - 0.3 - 0.4 = 0.3$ である. □

3.3　最適質問集合

最適質問集合を以下で定義しておく.

定義7　最適質問集合

ベイズ決定を行うときの平均決定誤り率を最小にする X を最適質問集合 X^{opt} と呼び, 次式で定義される.

$$X^{opt} = \arg \min_{X \in \mathcal{X}^K} BR_1(X, \hat{f}^{opt}_{X, \mathcal{Y}^K}, p(f)). \qquad (25)$$

□

補題3　最適質問集合は次式で与えられる．

$$X^{opt} = \arg \max_{X \in \mathcal{X}^K} \sum_{Y \in \mathcal{Y}^K} p(\hat{f}^{(1)}_{X,Y}). \quad \cdots\cdots\cdots\cdots \quad (26)$$

（証明）補題2より，明らか．　□

補題3は，$F_{X,Y}$ に含まれるブール関数の事前確率の最大値を全ての Y に対し和をとった値が最大となる X が，平均決定誤り率を最小とする X であることを意味している．

これまでに，各ブール関数に事前確率を付置した一般的な事前確率分布を対象とした場合[4]について最適質問集合を求める分枝限定アルゴリズムが提案されている[14]．しかし，アルゴリズムの計算量はブール変数の個数 n と質問回数 K の指数オーダであるため，n や K が大きい場合には実時間での計算は困難であった．

4．問題設定

この章では，事前確率分布に関する条件について説明した後，本論文で対象とする問題設定を説明する．未知である真のブール関数を $f(\underline{x}) = \Sigma_{\underline{a}} f_{\underline{a}} \mathcal{X}_{\underline{a}}(\underline{x})$ とする．

4.1　ブール関数の事前確率分布に関する条件

ここで，事前確率[5]であるが，一般には，何らかの基準で好ましいブール関数に高い事前確率を割り振ることが多い[2]．また，陽には仮定していないものの，ブール関数に優先順位を仮定した研究も事前確率分布を仮定しているととらえることができる．例えば，データに無矛盾なブール関数が複数個存在する場合，その中から１つを選ぶ最も標準的な方法は矛盾しない関数の中でその表現サイズが最小の関数を出力する方法である

[12]．これは決定理論から見た場合，表現サイズを基準とし，サイズの小さいものに高い事前確率を，大きいものに低い事前確率を割り振ることに相当する．また，実験計画法でも，より高次の成分を持つ関数を含むクラスほど重要度が低いと仮定されており[9]，これを決定理論から見ると，より高次の成分を持つ関数に小さい事前確率を割り振る事前確率分布を対象としているととらえることができる．

本節では，事前確率分布に関する条件を与える．まず必要な用語として，同じブール変数の集合によって関数値が決まるブール関数の集合を以下で定義する．

定義８　ブール関数集合 $F_{\underline{a}}$

ブール変数の集合を $S_{\underline{a}} = \{x_i | i \in v(\underline{a})\}$ とする．このとき，関数値が $S_{\underline{a}}$ に含まれるすべてのブール変数に依存し，かつ，これ以外のブール変数には依存しないブール関数の集合を $F_{\underline{a}}$ と定義する．

ここで，異なる二つのブール関数集合 $F_{\underline{a}}, F_{\underline{b}}, (\underline{a} \neq \underline{b})$ は互いに素，すなわち，$F_{\underline{a}} \cap F_{\underline{b}} = \phi$（$\phi$ は空集合）である．　□

$F_{\underline{a}}$ に含まれるブール関数の総数 $|F_{\underline{a}}|$ は次式で与えられる．

$$|F_{\underline{a}}| = 2^{2^{w(\underline{a})}} - \sum_{\underline{b} \in \{\underline{a}' | \underline{a}' \sqsubseteq \underline{a}\} \setminus \{\underline{a}\}} |F_{\underline{b}}|. \quad \cdots\cdots\cdots\cdots (27)$$

また，$f(\underline{x}) \in F_{\underline{a}}$ を直交基底により表現すると，次式で表現されることは明らかである．

$$f(\underline{x}) = \sum_{\underline{b} \in \{\underline{a}' | \underline{a}' \sqsubseteq \underline{a}\}} f_{\underline{b}} \mathcal{X}_{\underline{b}}(\underline{x}). \quad \cdots\cdots\cdots\cdots (28)$$

例５　F_{110} に含まれるブール関数の総数は 10 であり，それらをブール和∨を用いて表現[6]すると，

$$F_{110} = \{x_1x_2, \bar{x}_1x_2, x_1\bar{x}_2, \bar{x}_1\bar{x}_2, x_1x_2 \vee \bar{x}_1\bar{x}_2, \bar{x}_1x_2, \vee x_1\bar{x}_2,$$
$$x_1 \vee x_2, \bar{x}_1 \vee x_2, x_1 \vee \bar{x}_2, \bar{x}_1 \vee \bar{x}_2\}, \quad \cdots\cdots\cdots\cdots (29)$$

となる．また，これらのブール関数を基底関数の線形結合，すなわち，式(28)で表現すると，それぞれ以下となる．

$$
\begin{aligned}
x_1x_2 &\to -\frac{1}{2}\mathcal{X}_{000}(\underline{x}) - \frac{1}{2}\mathcal{X}_{100}(\underline{x}) - \frac{1}{2}\mathcal{X}_{010}(\underline{x}) + \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \\
\overline{x}_1x_2 &\to -\frac{1}{2}\mathcal{X}_{000}(\underline{x}) + \frac{1}{2}\mathcal{X}_{100}(\underline{x}) - \frac{1}{2}\mathcal{X}_{010}(\underline{x}) - \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \\
x_1\overline{x}_2 &\to -\frac{1}{2}\mathcal{X}_{000}(\underline{x}) - \frac{1}{2}\mathcal{X}_{100}(\underline{x}) + \frac{1}{2}\mathcal{X}_{010}(\underline{x}) - \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \\
\overline{x}_1\overline{x}_2 &\to -\frac{1}{2}\mathcal{X}_{000}(\underline{x}) + \frac{1}{2}\mathcal{X}_{100}(\underline{x}) + \frac{1}{2}\mathcal{X}_{010}(\underline{x}) + \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \\
x_1x_2 \vee \overline{x}_1\overline{x}_2 &\to \mathcal{X}_{110}(\underline{x}) \\
\overline{x}_1x_2 \vee x_1\overline{x}_2 &\to -\mathcal{X}_{110}(\underline{x}) \\
x_1 \vee x_2 &\to \frac{1}{2}\mathcal{X}_{000}(\underline{x}) - \frac{1}{2}\mathcal{X}_{100}(\underline{x}) - \frac{1}{2}\mathcal{X}_{010}(\underline{x}) - \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \\
\overline{x}_1 \vee x_2 &\to \frac{1}{2}\mathcal{X}_{000}(\underline{x}) + \frac{1}{2}\mathcal{X}_{100}(\underline{x}) - \frac{1}{2}\mathcal{X}_{010}(\underline{x}) + \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \\
x_1 \vee \overline{x}_2 &\to \frac{1}{2}\mathcal{X}_{000}(\underline{x}) - \frac{1}{2}\mathcal{X}_{100}(\underline{x}) + \frac{1}{2}\mathcal{X}_{010}(\underline{x}) + \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \\
\overline{x}_1 \vee \overline{x}_2 &\to \frac{1}{2}\mathcal{X}_{000}(\underline{x}) + \frac{1}{2}\mathcal{X}_{100}(\underline{x}) + \frac{1}{2}\mathcal{X}_{010}(\underline{x}) - \frac{1}{2}\mathcal{X}_{110}(\underline{x}) \quad \cdots\cdots\cdots (30)
\end{aligned}
$$

□

条件 1　事前確率分布に関する条件

1. $\forall f, f' \in F_{\underline{a}}$ に対し，$p(f) = p(f')$ が成立する．
2. ブール関数 $f \in F_{\underline{a}}$ の事前確率を $p(F_{\underline{a}})$ で表す．このとき，任意の $\underline{a}, \underline{b}, (\underline{a} \sqsubseteq \underline{b})$ に対し，$p(F_{\underline{a}}) \geq p(F_{\underline{b}})$ が成立する．□

条件 1 を満たす事前確率分布では，出力に影響を与えるブール変数の個数が大きいブール関数ほどそのブール関数の事前確率は小さくなる．

条件 1 は質問学習以外の分野でも見られる．例えば，実験計画法や統計解析では，より高次の成分を持つ関数を含むクラスほど重要度が低いと仮定することが通常である．実験計画法[9][7]において，重要度の高い構造式[8]は条件 1 で事前確率の高い構造式と等価である．

4.2 本論文で対象とする問題設定

アルゴリズムの入力について説明する．入力はブール関数の事前確率分布 $p(f)$ と質問回数 K, $(K = 2^k$, k は自然数) である．ただし，事前確率分布 $p(f)$ は条件1を満たすものとし，$m := |\{\underline{a} | p(F_{\underline{a}}) > 0\}|$ は n の多項式で抑えられるものとする．

ここで，$\{\underline{a} | p(F_{\underline{a}}) > 0\}$ に含まれる要素で $p(F_{\underline{a}})$ が i 番目に大きい要素 $\underline{a}$ を $\underline{a}^{(i)}$ で表す[9]．すなわち，$p(F_{\underline{a}^{(1)}}) \geq p(F_{\underline{a}^{(2)}}) \geq \cdots \geq p(F_{\underline{a}^{(m)}})$ である．これより，真のブール関数 f に対し，$f \in \cup_{i=1}^{m} F_{\underline{a}^{(i)}}$ が成立する．

アルゴリズムの流れは以下の通りである．アルゴリズムは，第1段階として，事前確率分布 $p(f)$ をもとに質問集合 X, $(|X| = K)$ を構成する．第2段階として，X に対し質問を行い，質問結果 $f(X)$ を得る．第3段階として，$X, f(X)$ に無矛盾で事前確率が最大なブール関数を構成し出力する．

5．質問回数が固定されているときの質問学習

この章では，質問集合に直交計画を用いたアルゴリズムを提案する．

5.1 必要な定義

アルゴリズムに必要な定義を与えておく．まず，$A_i = \{\underline{a} | \underline{a} \sqsubseteq \underline{a}^{(i)}\}$ とする．ここで，A_i が仮定1を満たすことは明らかである．以下で，ブール関数集合 F_A, F_{A_i} を定義しておく．

定義9　ブール関数集合 F_A, F_{A_i}

仮定1を満たす A, A_iに対し，ブール関数の集合 F_A を $F_A = \cup_{\underline{a} \in A} F_{\underline{a}}$, $F_{A_i} = \cup_{\underline{a} \in A_i} F_{\underline{a}}$ で定義する．　□

5.2 質問結果に無矛盾なブール関数がF_{A_i}に含まれる条件

X, $f(X)$ に無矛盾なブール関数が F_{A_i} に含まれる条件が以下で与えられことは明らかである．

条件２ $\underline{x} = (x_1, x_2, \cdots, x_n)$, $\underline{x}' = (x'_1, x'_2, \cdots x'_n)(\underline{x}, \underline{x}' \in X)$ が $\forall j \in v(\underline{a}^{(i)})$ において $x_j = x'_j$ であるとき，$f(\underline{x}) = f(\underline{x}')$ が成立する． □

例６ 条件２を満たす場合の例

$X = (0000, 1100, 1010, 0110, 1001, 0101, 0011, 1111)$, $f(X) = (+1, +1, +1, +1, +1, +1, -1, -1)$ であり，$\underline{a}^{(i)} = 0011$ である場合を考える．このとき，$f(0000) = f(1100)$, $f(1010) = f(0110)$, $f(1001) = f(0101)$, $f(0011) = f(1111)$, であるので，$\underline{a}^{(i)}$ は条件２を満たす．これより，X, $f(X)$ に無矛盾なブール関数が F_{A_i} に含まれることが分かる．また，そのブール関数は $\bar{x}_3 \vee \bar{x}_4$ である． □

5.3 アルゴリズム

質問結果に無矛盾で事前確率が最大なブール関数を求める自明なアルゴリズムとして，$\underline{a}^{(1)}$, $\underline{a}^{(2)}$, $\underline{a}^{(3)}$, $\cdots$ の順に条件２を満たすかどうかを調べ，最初に条件２を満たす $\underline{a}^{(i)}$ を $\underline{a}^{(t)}$ とすると，$F_{\underline{a}^{(t)}}$, $(F_{\underline{a}^{(t)}} \subseteq F_{A_t})$ に含まれるブール関数をしらみつぶしに探す方法が考えられる．ここで，$F_{\underline{a}^{(t)}}$ に含まれるブール関数の総数 $|F_{\underline{a}^{(t)}}|$ は K の指数オーダ[10]であるため，$F_{\underline{a}^{(t)}}$ に含まれるブール関数をしらみつぶしに探す計算量は K の指数オーダとなってしまう．

一方，質問集合が直交計画であれば定理１を用いることができ，F_{A_t} に含まれて質問結果に無矛盾なブール関数を構成する計算量は文献[15]より $O(K \log K)$ である．

そこで，直交計画を質問集合としたアルゴリズムを提案する．アルゴリズムの流れを以下に示す．

1．直交計画XAの構成

直交計画 X_A を出力するアルゴリズムを図１に載せる．このアルゴリズムについて説明する．質問回数が K であるため，まず大きさ K の直交計画が構成できる条件のもとで最大化した集合 A を求める．このとき，$A=\{\underline{a}^{(1)}, \underline{a}^{(2)}, \cdots, \underline{a}^{(|A|)}\}$ である．ここで，集合 A はブール関数の事前確率分布 $p(f)$ に依存する．そして，得られた集合 A に対し，直交計画 X_A を求めている．

```
入力：p(f), K　出力：X_A
i := 0; A := φ;
while G_A の行数 < log2K do
  begin
    i := i + 1; A := A∪a^(i);
    [13]のアルゴリズムを用い G_A を求める
  end;
G_A を用い，直交計画 X_A を構成する;
X_A を出力する.
```

図1：X_Aを出力するアルゴリズム

2．質問の実行

X_A を質問し，質問結果 $f(X_A)$ を得る．

3．ブール関数の出力

X_A, $f(X_A)$ に無矛盾で事前確率が最大なブール関数を出力する提案アルゴリズムを図２に載せる．

入力：$p(f)$, X_A, $f(X_A)$　出力：ブール関数 $\hat{f}_{X_A}$, $f(X_A)$

1：for $i := 1$ to m do

2：　begin

3：　　if $\underline{a}^{(i)}$ が条件２を満たす then

4：　　　begin

5：　　　　$A_i := \{\underline{a} \mid \underline{a} \sqsubseteq \underline{a}^{(i)}\}$;

6：　　　　$f_{\underline{a}} := \frac{1}{|X_A|} \Sigma_{\underline{x} \in X_A} f(\underline{x})\ \mathcal{X}_{\underline{a}}(\underline{x}), (\underline{a} \in A_i)$;

7：　　　　$\hat{f}_{X_A}, f(X_A) := \Sigma_{\underline{a} \in A_i} f_{\underline{a}} \mathcal{X}_{\underline{a}}(\underline{x})$;

8：　　　　$\hat{f}_{X_A}, f(X_A)$ を出力し，終了する

9：　　　end;

10：　end.

図2：提案アルゴリズム

5. 4　提案アルゴリズムの説明と計算量

提案アルゴリズムは，まず $\underline{a}^{(1)}, \underline{a}^{(2)}, \underline{a}^{(3)}, \cdots$ の順に条件２を満たすかどうかを調べ, 最初に条件２を満たす $\underline{a}^{(i)}$ を求める. この $\underline{a}^{(i)}$ を $\underline{a}^{(t)}$ とすると, 真のブール関数 f は $f \in \cup_{i=1}^{m} F_{\underline{a}^{(i)}}$ であるため，$t \leq m$ となる t が必ず存在する．$\underline{a}^{(t)}$ を見つけた後，定理１を用い $F_{\underline{a}^{(t)}}$ に含まれ質問結果に無矛盾なブール関数を構成する.

また，t が必ず存在することと $F_{\underline{a}^{(t)}}$ に含まれ質問結果に無矛盾なブール関数を構成可能であることより，アルゴリズムは必ず停止することが分かる.

次に，提案アルゴリズムの計算量について考える[11]．まず，３行目であるが，条件を満たすかどうか判定するには K 個の入力 $\underline{x}$ を並べ替えて $f(\underline{x})$ の値を比べればよい．このため，３行目で必要な計算量は $O(K \log K)$ である．６－７行目で必要な計算量は文献[15]より，$O(K\log K)$ である．また，４－９行目の処理は一度のみである．以上とfor文の繰り返しの回

数が m 以下であることより，提案アルゴリズムに必要な計算量は，$O(K\log K)$ となる．ここで，m は n の多項式で抑えられるため，提案アルゴリズムに必要な計算量は n と K の多項式オーダであることがわかる．

6．提案アルゴリズムの出力に関する最適性

この章では本論文の主結果である，提案アルゴリズムの出力に関する最適性を示す．

まず，必要な定義として，G_A の部分行列を以下で定義しておく．

$$G_A^{\underline{a}^{(i)}} = [\underline{g_{\cdot j1}}\ \ \underline{g_{\cdot j2}}\ \ \underline{g_{\cdot j_{|v(\underline{a}^{(i)})|}}}], \quad \cdots\cdots\cdots\cdots\cdots \quad (31)$$

ただし，$j_1, j_2, \cdots, j_{|v(\underline{a}^{(i)})|} \in v(\underline{a}^{(i)})$ である．

条件３

$\underline{r}, \underline{r}', (\underline{r}, \underline{r}' \in \{0, 1\}^{\log_2 K})$ が $\underline{r}G_A^{\underline{a}^{(i)}} = \underline{r}'G_A^{\underline{a}^{(i)}}$ であるとき，$f(\underline{r}G_A) = f(\underline{r}'G_A)$ が成立する． □

補題４　X が直交計画であるとき，条件２と条件３は等価である．

（証明）　直交計画の定義より明らか． □

補題５

$\underline{a}^{(i)}$ が条件２を満たすとき，$X_A, f(X_A)$ に無矛盾で F_{A_i} に含まれるブール関数の総数は $2^{2^{w(\underline{a}^{(i)})}}/2^{2^{rank(G_A^{\underline{a}^{(i)}})}}$ である．

（証明）　まず，F_{A_i} に含まれるブール関数の総数は $|F_{A_i}| = 2^{2^{w(\underline{a}^{(i)})}}$ である．一方，実質的な質問回数[12]は $2^{rank(G_A^{\underline{a}^{(i)}})}$ であり，質問結果は $2^{2^{rank(G_A^{\underline{a}^{(i)}})}}$ 通りである．以上より，補題が成立する． □

例7

$$G_A = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix}, \cdots\cdots\cdots\cdots\cdots\cdots \quad (32)$$

X_A = (00000, 11000, 10101, 01101, 10011, 01011, 00110, 11110), $f(X_A)$ = (+1, +1, +1, +1, +1, +1, −1, −1) であり，$\underline{a}^{(i)} = 00111$ A_i = {00000, 00100, 00010, 00001, 00110, 00101, 00011, 00111} である場合を考える．このとき，$f(00000) = f(11000)$, $f(10101) = f(01101)$, $f(10011) = f(01011)$, $f(00110) = f(11110)$, であるので，$\underline{a}^{(i)} = 00111$ は条件2を満たす．また，$w(00111) = 3$ であり，

$$G_A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \cdots\cdots\cdots\cdots\cdots\cdots \quad (33)$$

より $rank(G_A^{00111}) = 2$ である．以上より，X_A, $f(X_A)$ に無矛盾で $F_{A_i} = \cup_{\underline{a} \in A_i} F_{\underline{a}}$ に含まれるブール関数の総数は $2^{2^3}/2^{2^2} = 16$ である． □

補題6

$\underline{a}^{(i)}$ が条件2を満たす i の最小値を t とする．このとき，$rank(G_A^{\underline{a}^{(t)}}) = w(\underline{a}^{(t)})$ が成立する．

(証明) 背理法により，証明を行う．まず，$rank(G_A^{\underline{a}^{(t)}}) \leq w(\underline{a}^{(t)})$ が成立することは明らかである．そこで，$rank(G_A^{\underline{a}^{(t)}}) < w(\underline{a}^{(t)})$ が成立すると仮定する．この仮定より，

$$rank(G_A^{\underline{a}^{(s)}}) = rank(G_A^{\underline{a}^{(t)}}), \cdots\cdots\cdots\cdots\cdots\cdots \quad (34)$$

を満たす $\underline{a}^{(s)}$, ($\underline{a}^{(s)} \sqsubseteq \underline{a}^{(t)}$, $\underline{a}^{(s)} \neq \underline{a}^{(t)}$) が存在する．ここで，$\log_2 K \times$

$w(\underline{a}^{(s)})$ 行列 $G_A^{\underline{a}^{(s)}}$ は $\log_2 K \times w(\underline{a}^{(t)})$ 行列 $G_A^{\underline{a}^{(t)}}$ の部分行列である．このとき，$G_A^{\underline{a}^{(s)}}$ が $G_A^{\underline{a}^{(t)}}$ の部分行列であることと $rank(G_A^{\underline{a}^{(s)}}) = rank(G_A^{\underline{a}^{(t)}})$ より，$\underline{r}G_A^{\underline{a}^{(t)}} = \underline{r}'G_A^{\underline{a}^{(t)}}$ であれば $\underline{r}G_A^{\underline{a}^{(s)}} = \underline{r}'G_A^{\underline{a}^{(s)}}$ が成立する，すなわち，条件３が成立することは明らかである．以上と補題４より，$\underline{a}^{(s)}$ が条件２を満たすことがわかる．また，$\underline{a}^{(i)}$ の定義より $s < t$ であるため，t が条件２を満たす i の最小値であることに矛盾する．以上より，最初の仮定が誤りであることが分かり，補題が証明される．□

補題７

$\underline{a}^{(i)}$ が条件２を満たす i の最小値を t とする．このとき，F_{A_t} に含まれ，$X_A, f(X_A)$ に無矛盾なブール関数は唯一存在し，次式で与えられる．

$$\hat{f}_{X_A, f(X_A)} = \sum_{\underline{a} \in A_t} f_{\underline{a}} \mathcal{X}_{\underline{a}}(\underline{x}). \quad \cdots\cdots\cdots\cdots\cdots \quad (35)$$

ただし，係数 $f_{\underline{a}}$ は次式により計算される．

$$f_{\underline{a}} = \frac{1}{|X_A|} \sum_{\underline{x} \in X_A} f(\underline{x}) \mathcal{X}_{\underline{a}}(\underline{x}). \quad \cdots\cdots\cdots\cdots\cdots \quad (36)$$

（証明） F_{A_t} に含まれ，$X_A, f(X_A)$ に無矛盾なブール関数が唯一存在することは，補題５と補題６より明らかである．そこで，このブール関数が式(35)で与えられることを証明する．

まず，$rank(G_A^{\underline{a}^{(t)}}) = w(\underline{a}^{(t)})$ より，$\{\underline{g}_{\cdot j} | j \in v(\underline{a}^{(t)})\}$ が１次独立である．また，$\{\underline{g}_{\cdot j} | j \in v(\underline{a}^{(t)})\}$ が１次独立より，$\forall \underline{a}, \underline{b} \in A_t$ に対して $\{\underline{g}_{\cdot j} | j \in v(\underline{a}, \underline{b})\}$ が１次独立である．

以上より X_A は A_t の直交計画であることが分かる．これと定理１より，$X_A, f(X_A)$ に無矛盾なブール関数が式(35)で与えられることは明らかである．□

例８ $X_A, f(X_A)$ が例６の $X, f(X)$ であり，$\underline{a}^{(t)} = 0011$ の場合を考える．

このとき，X_A, $f(X_A)$ に無矛盾で F_{A_t} に含まれるブール関数 $\hat{f}_{X_A, f(X_A)}$ [13]は次式で与えられる．

$$\hat{f}_{X_A, f(X_A)} = \frac{1}{2}\mathcal{X}_{0000}(\underline{x}) + \frac{1}{2}\mathcal{X}_{0010}(\underline{x}) + \frac{1}{2}\mathcal{X}_{0001}(\underline{x}) - \frac{1}{2}\mathcal{X}_{0011}(\underline{x}) \quad \cdots\cdots\cdots\cdots (37)$$

ここで，係数は式(36)により計算される． □

定理２

ブール関数の事前確率分布が条件１を満たす場合，提案アルゴリズムはベイズ決定を行う．

(証明)　提案アルゴリズムが $\underline{a}^{(1)}$, $\underline{a}^{(2)}$, $\underline{a}^{(3)}$, $\cdots$ の順に条件２を満たすかどうかを調べることと補題７より，提案アルゴリズムで出力される $\hat{f}_{X_A, f(X_A)}$ は次式を満たす．

$$\hat{f}_{X_A, f(X_A)} = \hat{f}^{(1)}_{X_A, f(X_A)}. \quad \cdots\cdots\cdots\cdots\cdots\cdots (38)$$

補題１と式(38)より，この補題が証明される． □

７．提案アルゴリズムに関するその他の性質

この章では，直交計画が最適質問集合となる条件を与える．

7.1　平均決定誤り率の下界

平均決定誤り率の下界と最適質問集合の関係について説明する．事前確率 $p(f)$ が i 番目に大きい f を $f^{(i)}$ と表す[14]．すなわち，$p(f^{(1)}) \geq p(f^{(2)}) \geq p(f^{(3)}) \geq \cdots \geq p(f^{(|\mathcal{F}|)})$ である．このとき，式(23)より，平均決定誤り率の下界が以下で与えられることは明らかである．

$$BR_1(X,\ \hat{f}^{opt}_{X,\mathcal{Y}^K},\ p(f)) \geq 1 - \sum_{i=1}^{2^K} p(f^{(i)}). \quad \cdots\cdots\cdots\cdots \quad (39)$$

ここで$\forall f,\ f' \in \{f^{(1)},\ f^{(2)},\ f^{(3)},\ \cdots,\ f^{(2^K-1)},\ f^{(2^K)}\}$に対し，$f(X) \neq f'(X)$が成立する X が存在する場合，式(39)は等式で成立する．ただし，一般にはこのような X が存在するとは限らない．

また，本稿では質問方法が，質問回数 K の一括型を対象としているが，質問回数 K の逐次型（それまでの質問結果に依存して次に質問する入力を決定することができる）でも，平均決定誤り率の下界は式(39)で与えられる．

7.2 提案アルゴリズムを用いたときの平均決定誤り率

補題8

提案アルゴリズムを用いたときの平均決定誤り率は，次式で与えられる．

$$\begin{aligned} & BR_1(X_A,\ \hat{f}_{X_A},\ \mathcal{Y}^K,\ p(f)) \\ & = 1 - \sum_{i=1}^{|F_A|} p(f^{(i)}) - \sum_{Y \in \mathcal{Y}^K \setminus \mathcal{Y}^K_{F_A}} p(\hat{f}^{(1)}_{X_A,\ Y}). \quad \cdots\cdots\cdots \quad (40) \end{aligned}$$

ただし，$\mathcal{Y}^K_{F_A} = \{Y' \mid f(X_A) = Y',\ f \in F_A,\ Y' \in \mathcal{Y}^K\}$である．

（証明）　まず，式(23)と式(38)より，平均決定誤り率は次式で与えられる．

$$BR_1(X_A,\ \hat{f}_{X_A},\ \mathcal{Y}^K,\ p(f)) = 1 - \sum_{Y \in \mathcal{Y}^K} p(\hat{f}^{(1)}_{X_A,\ Y}). \quad \cdots\cdots \quad (41)$$

ここで，定理1より，任意の $f,\ f' \in F_A,\ (f \neq f')$ に対し，$f(X_A) \neq f'(X_A)$ が成立することは明らかである．これと $F_A = \{f^{(1)},\ f^{(2)},\ \cdots,\ f^{(|F_A|)}\}$ より，式(41)は以下のように書き換えることができる．

式(41)の右辺

$$= 1 - \sum_{Y \in \mathcal{Y}^K_{F_A}} p(\hat{f}^{(1)}_{X_A, Y}) - \sum_{Y \in \mathcal{Y}^K \setminus \mathcal{Y}^K_{F_A}} p(\hat{f}^{(1)}_{X_A, Y})$$

$$= 1 - \sum_{i=1}^{|F_A|} p(f^{(i)}) - \sum_{Y \in \mathcal{Y}^K \setminus \mathcal{Y}^K_{F_A}} p(\hat{f}^{(1)}_{X_A, Y}). \quad \cdots\cdots\cdots \quad (42)$$

□

7.3 直交計画が最適質問集合となる条件

直交計画が最適質問集合となる条件を与える前に，証明で必要となる補題を与えておく.

補題9

$\underline{a}^{(i)}$ が $rank(G_A^{\underline{a}^{(i)}}) = \log_2 K$ を満たすとき，任意の $f(X_A) \in \mathcal{Y}^K$ に対し，$\underline{a}^{(i)}$ は条件2を満たす.

(証明)　$rank(G_A^{\underline{a}^{(i)}}) = \log_2 K$ より，$G_A^{\underline{a}^{(i)}}$ の行ベクトル集合は1次独立である. このため，$\underline{r} \neq \underline{r}'$ であれば，$\underline{r}G_A^{\underline{a}^{(i)}} \neq \underline{r}'G_A^{\underline{a}^{(i)}}$ である. 以上より，任意の $f(X_A)$ に対し，$\underline{a}^{(i)}$ は条件2を満たす. □

例9　例7の G_A と $\underline{a}^{(i)} = 11100$ の場合を考える. $rank(G_A^{11100}) = 3$, $\log_2 K = 3$ であるため，補題9より，11100は条件2を満たすことが分かる. □

補題10

$rank(G_A^{\underline{a}^{(i)}}) = \log_2 K$ を満たす i, $(i \leq m)$ が存在する場合，この条件を満たす最小の i を i_{min} とする. このとき，任意の $f(X_A) \in \mathcal{Y}^K$ に対し，X_A, $f(X_A)$ に無矛盾なブール関数が $F_{A_{i_{min}}}$ に含まれる.

(証明)　補題9より，任意の $f(X_A) \in \mathcal{Y}^K$ に対し，$\underline{a}^{(i_{min})}$ は条件2を満たす. これより，補題が成立することは明らか. □

以上の準備のもとで，直交計画が最適質問集合となる条件は以下の定理で与えられる．

定理３　直交計画が最適質問集合となる条件

事前確率分布 $p(f)$が $p(F_{\underline{a}^{(i_{min})}}) = p(F_{\underline{a}^{(|A|+1)}})$ を満たすとき，次式が成立する．

$$X^{opt} = X_A. \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (43)$$

(証明)　補題10より，任意の $Y \in \mathcal{Y}^K$ に対して，X_A, Y に無矛盾なブール関数は $F_{A i_{min}}$ に含まれる．これより，任意の $Y \in \mathcal{Y}^K \diagdown \mathcal{Y}^K_{F_A}$ に対して，X_A, Y に無矛盾なブール関数は $\cup_{j=|A|+1}^{i_{min}} F_{\underline{a}^{(j)}}$ に含まれることがわかる．

一方，$p(F_{\underline{a}^{(i_{min})}}) = p(F_{\underline{a}^{(|A|+1)}})$ であるので，$\cup_{j=|A|+1}^{i_{min}} F_{\underline{a}^{(j)}}$ に含まれるブール関数の事前確率は全て等しく，この事前確率は $p(F_{\underline{a}^{(|A|+1)}})$ である．これより，任意の $Y \in \mathcal{Y}^K \diagdown \mathcal{Y}^K_{F_A}$ に対し，提案アルゴリズムが出力するブール関数の事前確率は $p(F_{\underline{a}^{(|A|+1)}})$ である．また，この事前確率が，$\mathcal{F} \diagdown F_A$ に含まれるブール関数の事前確率の最大値であることは明らかである．

以上より，$Y \in \mathcal{Y}^K \diagdown \mathcal{Y}^K_{F_A}$ ごとに提案アルゴリズムにより得られる $|\mathcal{Y}^K \diagdown \mathcal{Y}^K_{F_A}|$ 個のブール関数を $f^{(|F_A|+1)}, f^{(|F_A|+2)}, \dots, f^{(2^K)}$ としても一般性を失わない．これより，補題８を用いると平均決定誤り率は次式で与えられる．

$$BR_1(X_A, \hat{f}_{X_A}, \mathcal{Y}^K, p(f)) = 1 - \sum_{i=1}^{2^K} p(f^{(i)}). \quad \cdots\cdots\cdots \quad (44)$$

以上と式(39)より，直交計画 X_A は平均決定誤り率の下界を達成する，すなわち，最適質問集合 X^{opt} であることがわかる．　□

ここで，定理３の条件の意味と強さについて説明を行う．定理３の条件は，直交計画 X_A が $\forall f, f' \in \{f^{(1)}, f^{(2)}, f^{(3)}, \cdots, f^{(2^K-1)}, f^{(2^K)}\}$ に対し，$f(X_A) \neq f'(X_A)$ を満たすための条件を意味している．この条件は強い条件であるが，条件が成立する場合は直交計画さえ構成すれば最適質問集合を得ることができる．さらに，平均決定誤り率の下界である式(39)は質問方法が逐次型である場合でも成立するため，質問方法がより一般の質問回数 K の逐次型とした場合でも，定理３の条件を満たす場合，直交計画が最適質問集合となることが分かる．

最後に，定理３の条件を満たさない場合について考える．この場合，直交計画が最適質問集合となる保証は無いが，少なくとも $F_A = \{f^{(1)}, f^{(2)}, \cdots, f^{(|F_A|)}\}$ に関しては，任意の $f, f' \in F_A, (f \neq f')$ に対し，$f(X_A) \neq f'(X_A)$ が成立することは保証される．平均決定誤り率について，最適質問集合を用いた場合と直交計画を用いた場合の差を評価することは重要であるが，最適質問集合を求めるには全ての質問集合について平均決定誤り率を求めなければならない[15]．実時間で最適質問集合を求めることは計算量的に困難[16]であり，最適質問集合との差を評価することは非常に難しい問題である[17]．

8．アルゴリズムの実行例

この章では提案アルゴリズムの実行例を示す．

8.1　実行例の条件

ブール変数の個数 $n = 9$，質問回数は $K = 16$ とする．表１の事前確率分布 $p(f)$ が与えられた場合を考える．この事前確率分布は条件１を満たす．また，表１より $m = 64$ である．

図１のアルゴリズムにより得られた A と G_A を以下に載せる．

$$A = \{000000000, 100000000, 010000000, 110000000, 001000000, 101000000, 011000000, 000100000, 100100000, 010100000, 001100000, 000010000, 000001000, 000000100, 000000010, 000000001\}.$$

表1：事前確率分布$p(f)$

i	$\underline{a}^{(i)}$	$p(F\underline{a}^{(i)})$	i	$\underline{a}^{(i)}$	$p(F\underline{a}^{(i)})$
1	000000000	2.0×10^{-5}	33	001001000	1.4×10^{-5}
2	100000000	2.0×10^{-5}	34	001000100	1.4×10^{-5}
3	010000000	2.0×10^{-5}	35	001000010	1.4×10^{-5}
4	001000000	2.0×10^{-5}	36	001000001	1.4×10^{-5}
5	000100000	2.0×10^{-5}	37	000110000	1.4×10^{-5}
6	000010000	2.0×10^{-5}	38	000101000	1.4×10^{-5}
7	000001000	2.0×10^{-5}	39	000100100	1.4×10^{-5}
8	000000100	2.0×10^{-5}	40	000100010	1.4×10^{-5}
9	000000010	2.0×10^{-5}	41	000100001	1.3×10^{-5}
10	000000001	2.0×10^{-5}	42	000011000	1.3×10^{-5}
11	110000000	1.9×10^{-5}	43	000010100	1.3×10^{-5}
12	101000000	1.9×10^{-5}	44	000010010	1.3×10^{-5}
13	011000000	1.9×10^{-5}	45	000010001	1.3×10^{-5}
14	100100000	1.9×10^{-5}	46	000001100	1.3×10^{-5}
15	010100000	1.9×10^{-5}	47	000001010	1.3×10^{-5}
16	001100000	1.9×10^{-5}	48	000001001	1.3×10^{-5}
17	100010000	1.5×10^{-5}	49	000000110	1.3×10^{-5}
18	100001000	1.5×10^{-5}	50	000000101	1.3×10^{-5}
19	100000100	1.5×10^{-5}	51	000000011	1.3×10^{-5}
20	111000000	1.5×10^{-5}	52	110010000	0.5×10^{-5}
21	110100000	1.5×10^{-5}	53	110001000	0.5×10^{-5}
22	101100000	1.5×10^{-5}	54	110000100	0.5×10^{-5}
23	011100000	1.5×10^{-5}	55	110000010	0.5×10^{-5}
24	111100000	1.5×10^{-5}	56	110000001	0.5×10^{-5}
25	100000010	1.4×10^{-5}	57	101010000	0.5×10^{-5}
26	100000001	1.4×10^{-5}	58	101001000	0.5×10^{-5}
27	010010000	1.4×10^{-5}	59	101000100	0.5×10^{-5}
28	010001000	1.4×10^{-5}	60	101000010	0.5×10^{-5}
29	010000100	1.4×10^{-5}	61	101000001	0.5×10^{-5}
30	010000010	1.4×10^{-5}	62	011010000	0.5×10^{-5}
31	010000001	1.4×10^{-5}	63	011001000	0.5×10^{-5}
32	001010000	1.4×10^{-5}	64	011000100	0.5×10^{-5}

$$G_A = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}. \cdots\cdots\cdots\cdots \quad (45)$$

式(45)の G_A を用い構成される X_A を質問し，得られた質問結果は，
$f(000000000) = -1$, $f(110011000) = -1$, $f(101010100) = -1$,
$f(011001100) = -1$, $f(100110010) = -1$, $f(010101010) = -1$,
$f(001100110) = -1$, $f(111111110) = +1$, $f(011110001) = -1$,
$f(101101001) = -1$, $f(110100101) = -1$, $f(000111101) = -1$,
$f(111000011) = -1$, $f(001011011) = -1$, $f(010010111) = -1$,
$f(100001111) = -1$, であった．

8. 2 実行結果

条件2は，$i = 24$ のときに初めて成立した．このとき，得られた $\hat{f}_{X_A, f(X_A)}$ を以下に載せる．

$$\begin{aligned}
&\hat{f}_{X_A, f(X_A)}(\underline{x}) \\
&= -\frac{7}{8}\mathcal{X}_{000000000}(\underline{x}) - \frac{1}{8}\mathcal{X}_{100000000}(\underline{x}) \\
&\quad - \frac{1}{8}\mathcal{X}_{010000000}(\underline{x}) - \frac{1}{8}\mathcal{X}_{001000000}(\underline{x}) \\
&\quad - \frac{1}{8}\mathcal{X}_{000100000}(\underline{x}) - \frac{1}{8}\mathcal{X}_{011100000}(\underline{x}) \\
&\quad - \frac{1}{8}\mathcal{X}_{101100000}(\underline{x}) - \frac{1}{8}\mathcal{X}_{110100000}(\underline{x}) \\
&\quad - \frac{1}{8}\mathcal{X}_{111000000}(\underline{x}) + \frac{1}{8}\mathcal{X}_{111100000}(\underline{x}) \\
&\quad + \frac{1}{8}\mathcal{X}_{110000000}(\underline{x}) + \frac{1}{8}\mathcal{X}_{101000000}(\underline{x}) \\
&\quad + \frac{1}{8}\mathcal{X}_{011000000}(\underline{x}) + \frac{1}{8}\mathcal{X}_{100100000}(\underline{x}) \\
&\quad + \frac{1}{8}\mathcal{X}_{010100000}(\underline{x}) + \frac{1}{8}\mathcal{X}_{001100000}(\underline{x}). \cdots\cdots \quad (46)
\end{aligned}$$

次に，質問集合の最適性について考える．この実行例では，$i_{min} = 24$ であり，$p(F_{\underline{a}^{(i_{min})}}) = p(F_{\underline{a}^{(24)}}) = 1.5 \times 10^{-5}$ となる．一方，$|A| = 16$ であり，$p(F_{\underline{a}^{(|A|+1)}}) = p(F_{\underline{a}^{(17)}}) = 1.5 \times 10^{-5}$ となる．以上より，$p(F_{\underline{a}^{(i_{min})}}) = p(F_{\underline{a}^{(|A|+1)}})$ であるので定理3より，直交計画 X_A は最適質問集合 X^{opt} であることが分かる．また，このときの平均決定誤り率を求めると 0.01662 であった．

9．むすび

本論文では，質問する入力集合に直交計画を用いるアルゴリズムを提案した．そして，提案アルゴリズムは質問結果に対してベイズ決定を行うことを示し，直交計画が最適質問集合となる条件を与えた．今後の課題として，他の損失関数を用いた場合や確率的なブール関数に対しての質問学習アルゴリズムの構築等があげられる．

注釈

1 NP困難であっても実用的に意味のある規模があまり大きくない場合，利用可能である一般的なアルゴリズムとして分枝限定法が知られている．

2 すなわち，$0 = (0, 0, \cdots, 0)$ である．

3 G_Aを構成するアルゴリズムであるが，このアルゴリズム自体が1つの研究対象であり，これまで実験計画法の分野で多くの研究が行われている．直交計画の構成法 (G_Aの構成法) まで含めると論文の論点がぼけてしまうため，本論文では含めないこととした．

4 質問方法であるが，本稿の一括型と異なり，逐次型を対象としている．

5 事前確率の設定方法については本論文では触れない．詳しくは，文献[2]参照．

6 紙面の都合上，ここではブール和を用いた表現を載せている．

7 ただし，陽には構造式空間に事前確率分布は仮定していない．

8 質問学習のブール関数に対応する．

9 $p(F_{\underline{a}})$ が同じ $\underline{a}$ が複数存在する場合，その順番は任意で構わない．

10　$2^{w(\underline{a}^{(i)})}$ の最大値が K であることと式(27)より得られる.

11　(注3) でも述べたように, 図1の計算量については触れない.

12　質問集合 (直交計画) に対し, $v(\underline{a}^{(i)})$ に含まれる変数がとり得る値の組み合わせの総数である.

13　ブール和を用いて表現すると $\bar{x}_3 \vee \bar{x}_4$ である.

14　この表記は提案アルゴリズムの平均決定誤り率を説明するために用いたものであり, 平均決定誤り率は $p(f)$ が同じ f に対し順番 i をどのように決めるかに依存しないため, この順番は本質的な問題ではない.

15　ここで, もし $\forall f,\ f' \in \{f^{(1)}, f^{(2)}, f^{(3)}, \cdots, f^{(2^K-1)}, f^{(2^K)}\}$ に対し, $f(X) \neq f'(X)$ を満たす X が存在すればそれは最適質問集合であるが, 一般にこのような X が存在するとは限らない.

16　質問集合の総数は K の指数オーダである.

17　n, K が小さい場合に限り, 分枝限定法等を用いることで最適質問集合を得ることができる[14].

参考文献

[1] D. Angluin, "Queries and concept learning," Machine Learning, Vol.2, No.4, pp.319-342, 1988.

[2] J. O. Berger, Statistical Decision Theory and Bayesian Analysis, Springer, 1985.

[3] T.S.Ferguson, Mathematical Statistics, Academic Press, 1967.

[4] E. M. Gold, "Language identification in the limit," Inf. Control, Vol.10, pp.447-474, 1967.

[5] D. Haussler, "Decision Theoretic Generalizations of the PAC Learning Model," 1st International Workshop, ALT'90, pp.21-41, 1990.

[6] E. Kushilevitz and Y. Mansour, "Learning Decision Trees using the Fourier Spectrum," SIAM J. Comput., Vol.22, No.6, pp.1331-1348, 1993.

[7] N. Lineal, Y. Mansour, and N. Nisan, "Constant depth circuits, Fourier Transform and learnability," J. ACM., Vol.40, No.3, pp.607-620, 1993.

[8] 松嶋敏泰, "帰納・演繹推論と予測-決定理論による学習モデル-, 1998年情報論的学習理論ワークショップ予稿集, pp.1-8, 1998.

[9] 奥野忠一, 芳賀敏郎, 実験計画法, 培風館, 東京, 1969.

[10] G. Paass.(著), 麻生英樹(訳), "予測とモデル選択のための質問選択," 情報処理, Vol.38, No.7, pp.562-568, Nov. 1997.

[11] 佐川雅彦, 貴家仁志, 高速フーリエ変換とその応用, 昭晃堂, 1992.

[12] 榊原康文, 小林聡, 横森貴, 計算論的学習, 培風館, 東京, 2001.

[13] 須田健二，宮崎晴夫，"直交配列を用いた実験計画における要因割りつけのコンピュータ・アルゴリズムについて，" 日本経営工学会誌, Vol.37, No.6, pp.345-352, 1987.

[14] 浮田善文，松嶋敏泰，平澤茂一，"質問からの学習問題の決定理論による定式化に関する一考察，" 情報処理学会論文誌，Vol.39, No.11, pp.2937-2948, Nov. 1998.

[15] 浮田善文，松嶋敏泰，平澤茂一，"直交計画を用いたブール関数の学習に関する一考察，" 電子情報通信学会論文誌，Vol.J86-A, No.4, pp.482-490, Apr. 2003.

[16] Y. Ukita, T. Saito, T. Matsushima and S. Hirasawa, "A Note on a Sampling Theorem for Functions over GF(g) ^{n} Domain,"IEICE Trans. Fundamentals, Vol.E93-A, No.6, pp.1024-1031,June 2010.

[17] Y. Ukita and T. Matsushima, "A Note on Estimation of theEffects in the Experimental Design using Fourier Transforms," in *Proc. The 2010 IEICE Society Conference*, A-4-4, p.73, Sep. 2010.

[18] Y. Ukita, T. Matsushima and S. Hirasawa, "Estimation of the Effects in the Experimental Design using Fourier Transforms," IEICE *Trans. Fundamentals,* Vol.E93-A, No.11, pp.2077-2082, Nov. 2010.

[19] L. G. Valiant, "A theory of the learnable," Commun. ACM, Vol.27, pp.1134-1142, 1984.